Department of Industrial Systems Engineering and Management | System Design Project

Development of Deep Learning Models for Abnormality Detection in Small Bowel Capsule Endoscopy Images

Supervised by A/P CHEUNG Wang Chi

Lim Zhan Yi Zachary, Koh Tze Yee, Lim Wei Sheng Dexter, Zhang Licheng, Zhang Kexuan

Executive Summary

Our group project is titled: Development of Deep Learning Models for Abnormality Detection in Small Bowel Capsule Endoscopy Images. The goal of this project is to develop a model for SingHealth to serve as a decision support system for detection of abnormalities when diagnosing small bowel capsule endoscopy images. The project will focus on the building and training of the detection models required for the diagnosis process.

Problem Description

Background

- The endoscopy is a medical procedure that allows doctors to examine the digestive tract using an instrument called an endoscope.
- Capsule endoscopy involves patients ingesting a small pill camera that captures 50,000-60,000 photos of the digestive tract.
- The examination of 50,000 pictures per patient is laborious, ranging from 45 minutes to 3 hours for doctors, increasing the chance of missing abnormalities.

Objectives

- Develop a model to serve as a decision support system for detection of abnormalities when diagnosing capsule endoscopy images.
- Reduce the amount of endoscopy images.

Partner Company

Singapore Health Services (SingHealth) is the largest organization of healthcare institutions that offer healthcare that is accessible, inexpensive, and of high quality. It provides comprehensive, interdisciplinary, and integrated care with over 40 clinical specializations, a network of acute hospitals, national specialty centers, polyclinics, and community hospitals.

Methodologies

Data

- Kvasir-Capsule dataset and SingHealth dataset with 32,524 and 36,171 images respectively. Each dataset was manually vetted, and labels were verified for accuracy.
- Data class labels are skewed towards "Usable" and "Normal" sets. To handle this imbalance, we utilize the Albumentations Python package for image augmentation to resample images in the minority classes, thereby creating new synthetic data.
- Kvasir-Capsule dataset is split into 80% training, 20% testing; SingHealth Dataset is split into 50% tuning, 50% validation.

Methods

Convolutional Neural Network (CNN):

- CNN utilize convolutional filters to extract features from images.
- Two CNN models are built using the TensorFlow package:
- **Bowel prep model (BPM)** filters usable and unusable images
- **Abnormality detection model (ADM)** detects the presence of abnormalities in photos
- **Transfer learning** is utilised to leverage on pre-trained models as feature extraction modules. Features are passed on to additional layers to further extract CE diagnosis-specific features for more accurate classification.

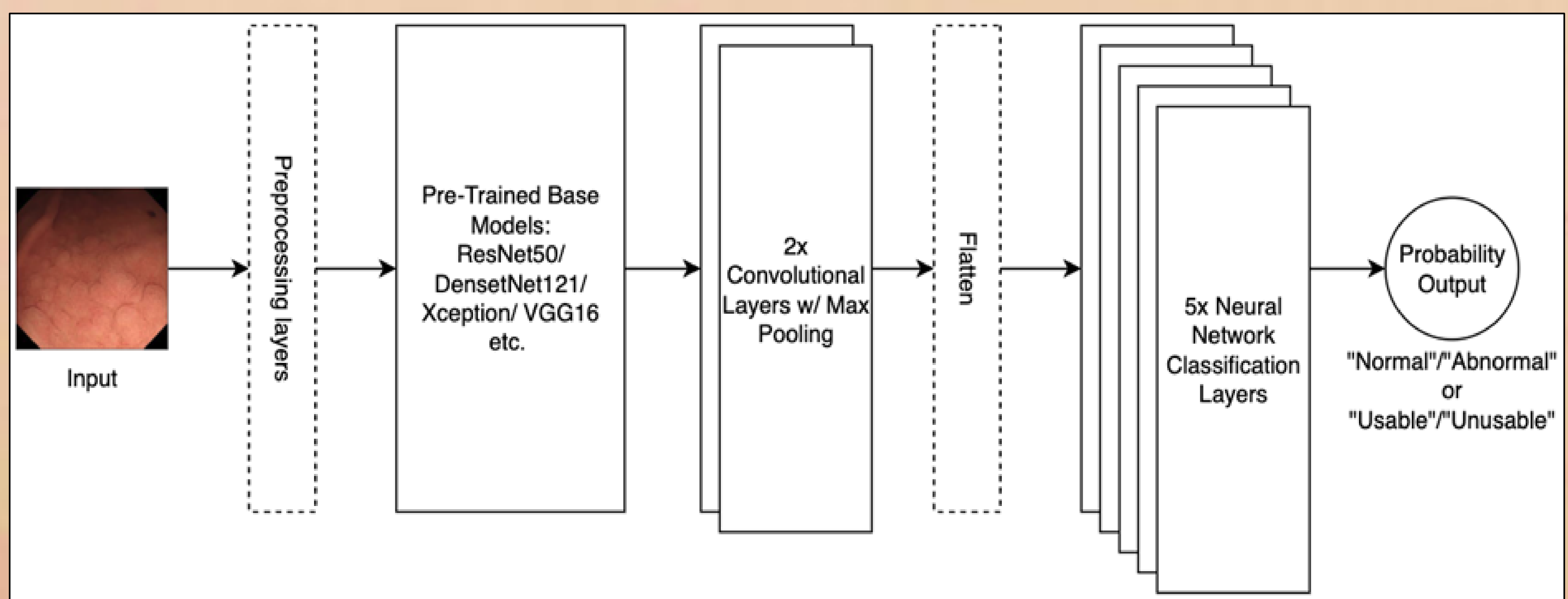


Figure 1: Transfer Learning Model Architecture

Technical Skillsets

The team had the opportunity to apply **machine learning methods**:

- **Image augmentation** techniques to process data
- **TensorFlow framework** to build deep learning models
- **Quality metrics** to evaluate model performance
- Implementation of drivers and packages on Linux systems

The team also learnt the inner workings of deep learning convolutional neural networks and its advantages in computer vision applications.

Results & Application

Performance Metrics

Models are benchmarked on their predictive performance on an out-of-sample validation set. We utilize metrics that are robust against data imbalance. The classification threshold is tuned to maximize F1-Score and we also evaluate the AU-PRC score. Models are selected based on F1-Score in order to strike a balance between precision and recall.

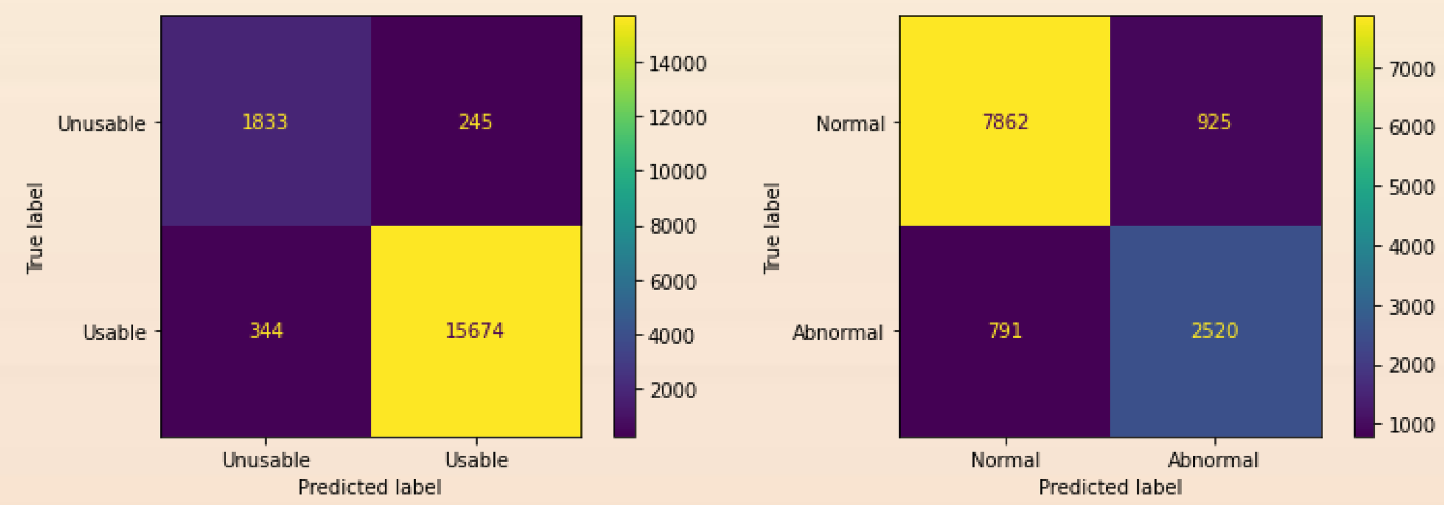
SingHealth Validation Dataset

Model Name	Precision	Recall	F1-Score	AU-PRC
BPM Random Forest (no tuning)	0.9840	0.0576	0.1089	0.9207
BPM CNN	0.9879	0.0281	0.0546	0.8637
BPM VGG16 base w/ CNN	0.9846	0.9791	0.9819	0.9984
ADM Xception base w/ CNN	0.6923	0.7834	0.7351	0.8266
ADM DenseNet121 base w/ CNN	0.7315	0.7611	0.7460	0.8415

Models Chosen

BPM : VGG16 CNN as base model, pretrained on ImageNet dataset

ADM : DenseNet121 CNN as base model, pretrained on ImageNet dataset



BPM VGG16 base w/ CNN **ADM DenseNet121 base w/ CNN**

Figure 2: Confusion Matrix for selected BPM and ADM models

From the Gain chart, we can observe that 43.2% of abnormalities are captured in the first decile. This represents a lift value of 8.64, where lift represents the ratio of observing an abnormality in the decile with the model as compared to observing an abnormality without the model

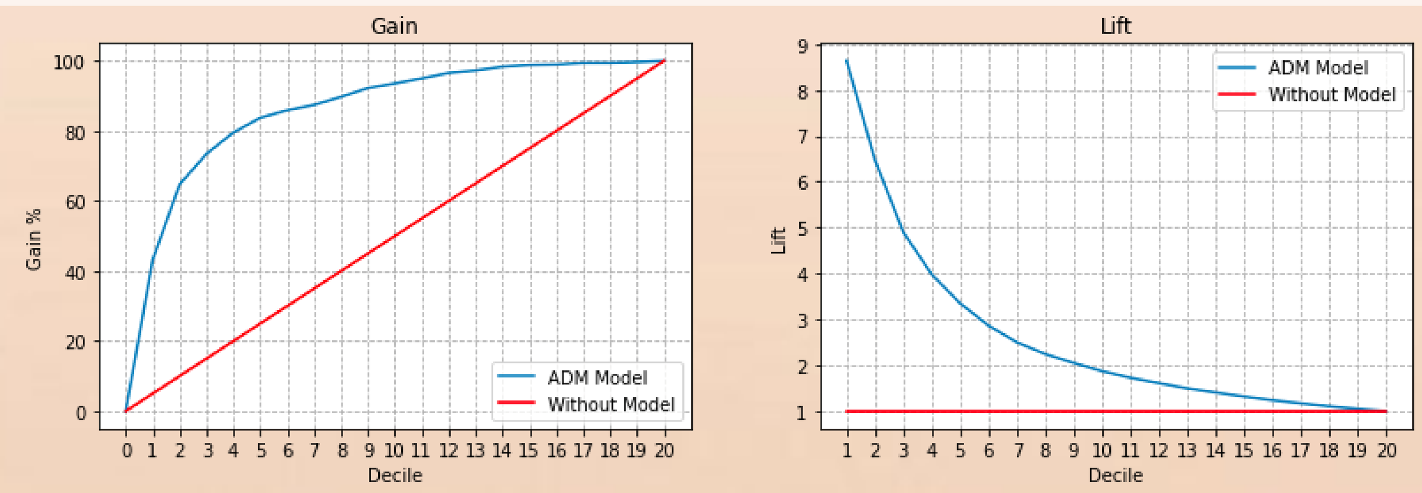


Figure 3: Lift and Gain Charts for ADM model

Implementation - Diagnosis Assistant Framework

- The diagnosis assistant framework breaks down the process into 2 stages: filtering out usable images; labelling images with abnormalities.
- A web-based application implements a user-friendly interface for uploading and analyzing images

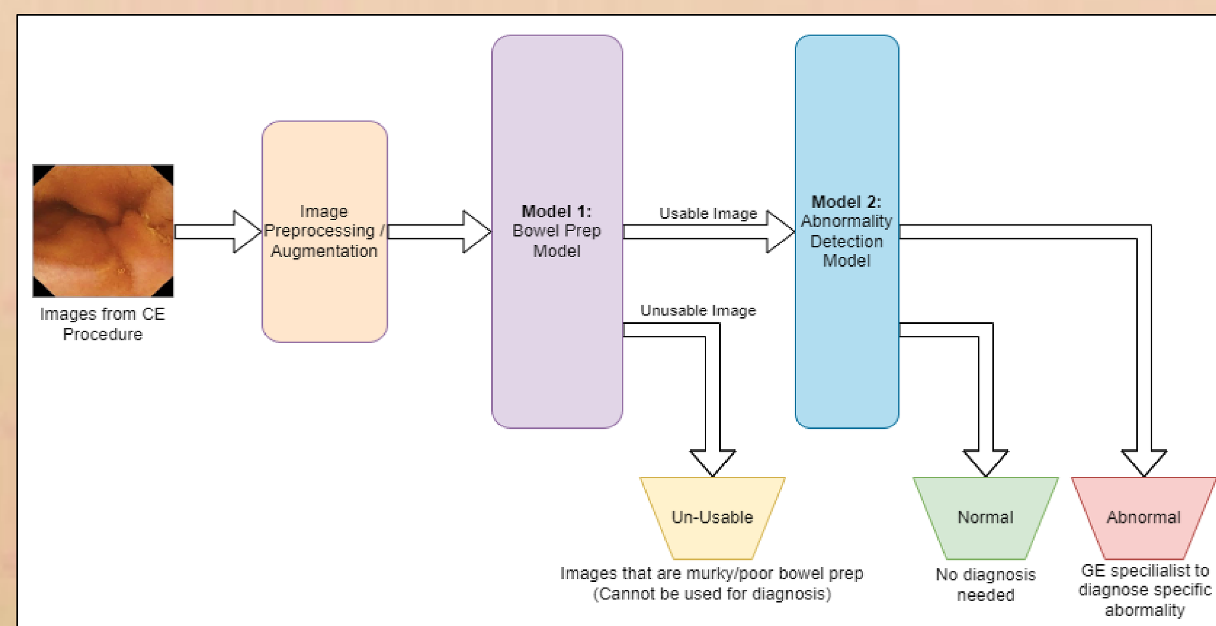


Figure 4: Proposed Diagnosis Assistant Framework

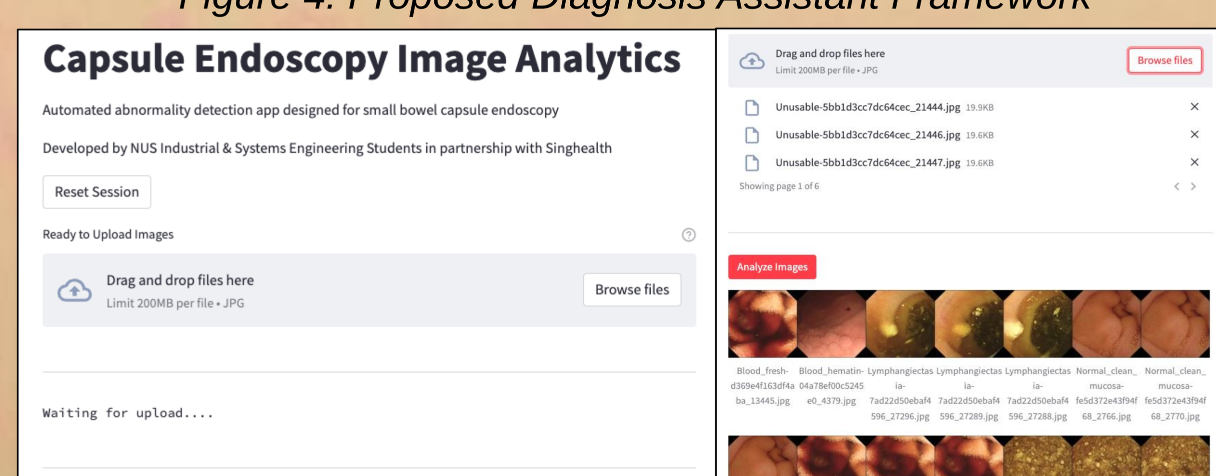


Figure 5: User Interface in Streamlit application